

## **EVALUATING SIMULATORS FOR USE IN ASSESSMENT OF MARINER PROFICIENCIES<sup>1</sup>**

*Dr. Marvin C. McCallum, Battelle Seattle Research Center, Seattle, WA*  
*Dr. Myriam Witkin Smith, U.S. Coast Guard Research & Development Center, Groton, CT<sup>2</sup>*

### **Abstract**

The paper describes a study sponsored by the U.S. Coast Guard to examine the implications of the STCW mandate for assessment of mariner proficiency by practical demonstration before a qualified assessor. Among the products of the study was a structured approach to evaluating simulators in their capability to support mariner assessment. The approach was refined by the trial evaluation of two desktop simulators for the assessment of mariner proficiency in ARPA operation.

### **Introduction**

The assessment of mariner proficiencies by practical demonstration, as mandated by the International Maritime Organization (IMO) in its 1995 Amendments to the Seafarer's Training, Certification and Watchkeeping (STCW) Code, is a substantial departure from earlier practice. Therefore, methods for developing, implementing, and documenting such assessments need to be established. The United States Coast Guard sponsored a research project to examine the implications of the mandate and to ensure that the best practices of current technology were available to the maritime industry. A group consisting of researchers, faculty from several maritime academies, and a commercial shipping company developed a model approach to developing assessments that are technically rigorous, and then demonstrated the practicality of the approach by implementing sample assessments on simulators, training vessels, and, finally, commercial vessels. Project efforts are reported elsewhere, including in IMO MSC Circular 853.

### **An Approach to Developing Performance-based Assessments**

Our earliest efforts on this project were to design a systematic method for the development of performance-based assessments of mariner proficiencies. Our project team examined the IMO STCW Code requirements, the methods of Instructional Systems Development (ISD), and the best practices of the maritime industries. Our intention was to distill a method that preserved the ISD objectives of establishing valid and reliable assessment objectives, measures, and standards while streamlining the process to serve the present needs of the maritime industry. Our five-step method is illustrated in

---

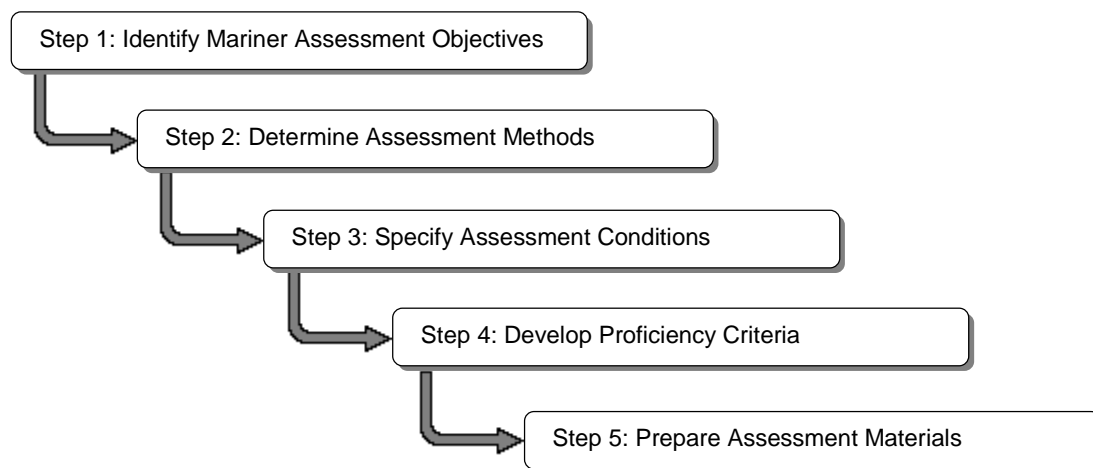
<sup>1</sup> The views expressed in this paper are those of the authors and are not official U.S. Coast Guard policy.

<sup>2</sup> The authors of this paper wish to gratefully acknowledge the participation of a varied and capable team, including especially Ms. Mireille Raby, now of the University of Iowa; Ms. Alice M. Forsythe, of Battelle Seattle Research Center; and Captains John M. Nunnenkamp and George R. Sandberg, both of the U.S. Merchant Marine Academy.

Figure 1. Step 1 is to *identify assessment objectives*, the critical elements of performance that will be assessed. In the language of STCW, these are “knowledge, understanding, and proficiencies.” A valid list is achieved by consulting the STCW Code, other relevant regulations, job procedures, task analyses, and subject matter experts. At this point in the process, the list of mariner assessment objectives could be customized to include requirements to follow the job procedures of a specific shipping company or to work with the equipment of a specific manufacturer.

With the assessment objectives identified, it is possible to go on to Step 2 and *determine assessment methods* for each performance element, whether as written or oral questions, exercises in a simulated job setting, or exercises in an actual job setting. Step 3 is to *specify assessment conditions*, a more detailed planning of how the assessment will be conducted. Step 4, to *develop proficiency criteria*, is the least intuitive part of the process, but extremely critical to the validity and reliability of assessments. Criteria are based on performance measures and standards. A “performance measure” is an element of the required activities that can be observed or measured; a “performance standard” is the acceptable or target level of an individual performance measure that is required for proficiency. An example of an objective might be to assess a mariner’s ability to steer by gyrocompass. The corresponding observable measure might be the accuracy with which the mariner is able to maintain the ordered heading. The standard, in this case, might be to maintain the ordered heading to an accuracy of plus or minus three degrees. To continue this example, if the mariner can achieve that accuracy of steering under specified operational conditions, he/she is to be considered proficient in steering by gyrocompass.

Step 5 is to *prepare assessment materials*, a final package of forms and documentation that should be adequate to guide the candidate and the assessor in the conduct of the assessment. An overview of the method is given in IMO, 1998; the design of the method is described in McCallum, Forsythe, Nunnenkamp, Sandberg, and Smith, 1999; and guidance for its use is given in McCallum, Forsythe, Barnes, Smith, Macaulay, Sandberg, Murphy, and Jackson, 1999.



**Figure 1.** Five-Step Method for Developing Mariner Performance-Based Assessments

As we designed the method, we tested and refined it with a case study, an assessment of mariner proficiency in Automatic Radar Plotting Aid (ARPA) operation. STCW mandates that this competence be assessed on a simulator and provides considerable detail and guidance as to the assessment objectives. We derived our assessment objectives from an analysis of ARPA proficiencies

identified in the STCW Code, then concentrated our efforts on the details of the assessment conditions and the criteria. We were fortunate to have faculty at the U.S. Merchant Marine Academy (USMMA) as partners in this effort, individuals with extensive expertise in the training and assessment of mariner skills and in the use of simulators for that purpose. USMMA has a fully-capable ARPA laboratory with simulators based on real equipment, so we were able to develop our mariner assessment with little concern for any limitations of the simulator.

A sample excerpt from our ARPA assessment procedure appears in Table 1, to illustrate the method. The two mariner assessment objectives listed here come directly from the STCW Code, Section B-I/12, *Guidance regarding the use of simulators*, paragraphs 18 through 35, *Training and assessment in the operational use of automatic radar plotting aids (ARPA)*. Assessment Objective 2.1 requires a demonstration of the mariner's understanding of the criteria for selection of targets by automatic acquisition. With our USMMA partners, we developed simulator exercises that required the candidate to demonstrate the performance of interest, in this case, setting up the ARPA with guard zones. Performance measures were also defined that specified the outcome of mariner performance was to be observed in order to assess performance. For Objective 2.1, the guard zone settings made by the mariner constituted the performance measure. The complete ARPA assessment, including descriptions of the simulator exercises, is available for review and for adoption in our first report (McCallum, Forsythe, Nunnenkamp, Sandberg, and Smith, 1999).

**Table 1.** Sample Mariner Assessment Objectives – ARPA Use for Situation Assessment

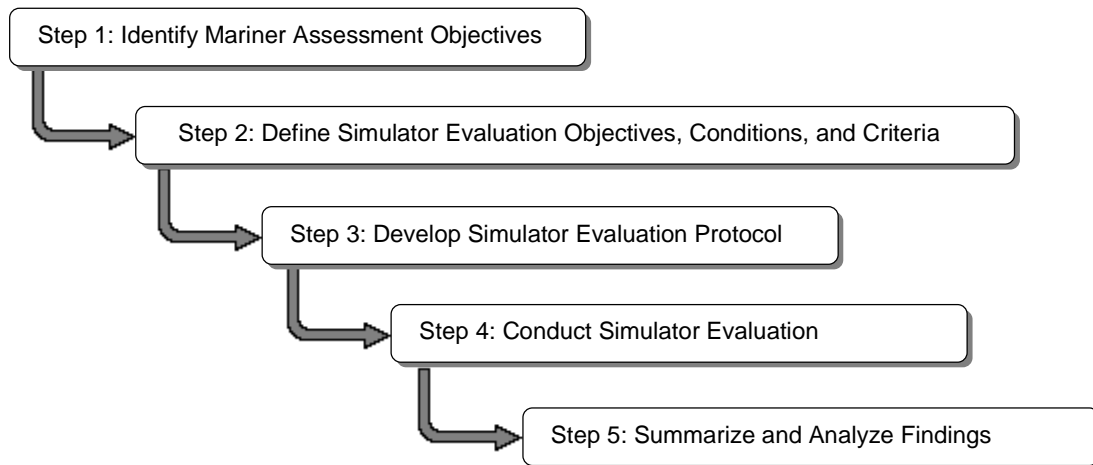
<b>Mariner Assessment Objective</b>	<b>STCW Reference for Objective</b>	<b>Assessment Condition</b>	<b>Performance Measures</b>
2.1 Understanding criteria for selection of targets by automatic acquisition	Section B-I/12, paragraph 25.1	Exercises A & B ARPA set-up instructions for setting guard zones	2.1.1 Guard zone settings
2.2 Appreciation of uses, benefits, and limitations of ARPA operational warning	Section B-I/12, paragraph 27.0	Exercise B ARPA set-up instructions for setting safe limit	2.2.1 Safe limit settings 2.2.2 Safe limit warning 2.2.3 Guard zone warning

## **An Approach to Evaluating Simulators**

At the present time, a broad range of marine simulators are available to support any given mariner assessment. Simulators range from those designed to provide highly realistic operational environments, controls, and displays to those that are designed to represent a facsimile of only some limited portion of the operational situation. Obviously, a simulator that is a full-scale mock-up of the operational setting is more likely to have many of the features and functions necessary to conduct a valid mariner assessment. However, with the advances in personal computer (PC) processing capability and the incorporation of PC processing in shipboard equipment, the advantages of the more complex and elaborate simulators are becoming less pronounced. When these advances are considered in conjunction with the affordability of PC-based simulators, it is evident that the capability of PC-based simulators to support mariner assessment must be seriously considered. Thus, another

major objective of our investigation was to design and test a systematic method to evaluate the capability of a simulator to support a specified set of assessment objectives.

Our method is illustrated in Figure 2. Step 1 is to *identify mariner assessment objectives*. We assume that these are not developed solely for the purpose of evaluating the simulator but must be developed for the larger purpose of assessing (or training) mariner performance. We designed and tested our general approach to simulator evaluation by continuing our ARPA case study, using the mariner assessment objectives that we had already developed. Step 2 in Figure 2 is the specification of the *simulator evaluation objectives, conditions, and criteria*, which establish the basis for determining if specific systems can support the assessment objectives. In specifying evaluation objectives, we considered, first, the simulator features needed to demonstrate the mariner performance to be assessed. We also considered additional simulator requirements described in STCW in Section A-I/12, *Standards governing the use of simulators*, Part 1, *Performance standards*, paragraph 2, *General performance standards for simulators used in assessment of competence*. These comprise the simulator's ability to simulate the capabilities of real equipment, to provide behavioral realism, to provide an interface to the mariner, to provide a variety of operating conditions, and to permit an assessor to control, monitor, and record exercises. We consulted IMO standards for the capabilities of real ARPA and navigational radar (IMO, 1979, 1971). For each simulator feature required to support the assessment objectives, we defined the conditions under which it would be evaluated, including the tasks that the mariner would perform and any environmental effects that would influence the performance of the equipment. Criteria are based on the simulator's ability to provide the necessary controls or displays for the mariner's tasks, or the assessor's needs, or for a simulation of real equipment. More discussion of how to develop the objectives, conditions, and criteria, and our complete ARPA example are available in a project report (Raby, Forsythe, McCallum, and Smith, 1999).



**Figure 2.** Method for Evaluating Simulators

To illustrate Step 2, a summary of two simulator evaluation objectives is presented in Table 2, which continues the example from the preceding table. Simulator Objective 3.2 requires the simulator to allow the use of automatic acquisition of targets. This capability of the simulator is necessary for Mariner Objective 2.1, the demonstration of an understanding of the criteria for the selection of targets by automatic acquisition. The necessary traffic and environmental conditions for the simulator

evaluation are those included in the exercises prepared earlier for the mariner assessment. Specific simulator capabilities that supported the control and display of targets through manual and automatic acquisition were identified and defined in terms of evaluation criteria. The specific criteria used to evaluate the capabilities of the simulators for Objective 3.2 consisted of four separate control-related criteria (e.g., *ability to manually acquire, track, process, and continuously update information for at least 10 targets*) and five separate display-related criteria (e.g., *ability to simultaneously display information for at least 10 targets in manual mode*). For the two example simulator objectives in Table 2, a total of 24 separate simulator evaluation criteria were defined.

**Table 2.** Sample Simulator Evaluation Objectives – ARPA Use for Situation Assessment

<b>Simulator Objective</b>	<b>Mariner Assessment Objective</b>	<b>Evaluation Condition</b>	<b>Simulator Control and Display Evaluation Criteria</b>
3.2 Use of manual and automatic acquisition	2.1 Understanding criteria for selection of targets by automatic acquisition	Acquire and track one or more targets using manual acquisition and using automatic acquisition	Control target acquisition in manual and automatic modes (four criteria)  Display targets and indicate acquisition mode (five criteria)
3.3 Use and limitations of ARPA operational warning	2.2 Appreciation of uses, benefits, and limitations of ARPA operational warning	Target violates safe limit area and activates warning	Control specific operational warning modes and parameters (six criteria)  Display operating mode and provide visual and/or audible warnings (nine criteria)

Step 3 is to develop a *simulator evaluation protocol*, a package of forms and documentation that will support the conduct of the evaluation of a candidate simulator. The protocol should describe, for each objective, the evaluation conditions needed to ensure comparability across evaluators or across simulators and the evaluation criteria for controls and displays. It should include a means of rating how well the feature satisfied the evaluation criteria. An excerpt from the protocol, illustrating sample evaluation criteria, is presented in Table 3. The table shows a sample of two control-related criteria and two display-related criteria for Objective 3.2 (from Table 2). Evaluators were instructed to rate the extent that each criterion was met (Yes, No, Partial) and provide brief comments regarding their ratings. A discussion of ways to organize a protocol and our complete ARPA protocol are presented in Raby, et al. (1999). Step 4, to *conduct the simulator evaluation*, and Step 5, to *summarize and analyze findings*, are described along with an overview of our case study in the next section.

**Table 3.** Sample Excerpts from Simulator Evaluation Protocol

Simulator Evaluation Objective	Simulator Evaluation Criterion	Reference	Rating			Comments
			Yes	No	Partial	
3.2 Use of manual and automatic acquisition	3.2.Control-1 Ability to acquire, track, process, and continuously update information for at least 10 targets	STCW A.422	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
	3.2.Control-2 Ability to acquire, track, process, and continuously update information for at least 20 targets	STCW A.422	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
	3.2.Display-1.1 Ability to display information for at least 10 targets simultaneously	STCW A.422	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
	3.2.Display-1.2 Indication of tracked targets	STCW A.422	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	

### Trial Application of Protocol and Analysis of Results

To test the approach, we conducted an evaluation of two commercial off-the-shelf PC-based ARPA simulators. Because cost is such an important factor in selecting simulators, we selected simulators that differed considerably in complexity and cost. Simulator *X* was of moderate complexity and cost; Simulator *Y* more closely approximated real equipment and was more costly. We were fortunate in having the assistance of the manufacturers in both cases, ensuring that each simulator exhibited its best potential. We forwarded to each manufacturer an overview of the evaluation objectives and a detailed description of the exercise scenarios. Subsequently, we sent a team of three researchers to apply the protocol and perform independent ratings. Our ratings for each criterion indicated how well the subject simulator met, partially met, or did not meet that criterion. We later combined the ratings for the researchers to achieve one consensus rating for each criterion. Even without further analysis, a comprehensive listing of each simulator criterion with its rating serves to identify the strengths and weaknesses of a simulator and provides detailed information regarding its effective use and potential areas for its improvement by the manufacturer.

We explored the potential use of these ratings further by performing a number of more complex analyses. First, we examined each simulator's capability to support mariner assessment. For each simulator, we calculated summary scores representing its capability to support each of six categories of mariner assessment objectives: setting up and maintaining displays, situation assessment, factors affecting performance and accuracy, parallel indexing, application of COLREGS (Collision Regulations), and operational warnings and system tests. Ratings of the simulator features needed to support the mariner assessment objectives in Table 1 are included in the summary score for situation assessment. We found that the more costly simulator was generally more capable, supporting more of the mariner assessment objectives, but neither of the PC-based simulators was fully capable of supporting all of the objectives. We also examined each simulator's capability to provide the general simulator capabilities taken from STCW, by grouping the individual criteria into the categories of exercise programming, equipment set-up, simulation, and debriefing. Again we found that the more

costly simulator better met those requirements but neither simulator met all of the requirements. Our analysis approach is discussed in a project report (Raby, Forsythe, McCallum, and Smith, 1999).

## Conclusions

We designed, tested, and demonstrated the technical feasibility of rigorous evaluation of marine simulators to support carefully developed mariner assessments. Our approach is fully compliant with the STCW Code's performance standards for simulators supporting mariner assessments. However, we intended that our approach be widely adapted and generalized. We believe that the approach can be useful to a wide variety of potential users. The detailed evaluation method that we have presented here is a tool not only for the assessment developer, but also for the simulator designer. The evaluation identifies weak features and the potential value of their improvement to the user. After identifying weaknesses, the evaluator can consider the value of a potential improvement in relation to its cost to the manufacturer and to future buyers. We believe that our approach can be readily adapted to evaluating simulators for training. The requirements for training system design, that the performance objectives and performance measures be carefully specified, are similar to those for assessment. Finally, we believe that the approach can be applied to many types of simulators. We would include those representing a variety of ship bridge and engine room equipment; those representing other maritime applications like vessel loading and vessel traffic systems; and even those meant for other applications like flight, driving, and power plant control simulators.

Despite our finding that specific simulators did not fully support specific mariner assessments, the potential advantages of using simulators rather than real equipment, or simulators based on real equipment, remain. We feel that the question to be asked is not, "Can simulators support assessment of mariner proficiency?" Instead, the question should be, "What is the best use of simulators in assessment of mariner proficiency?" A simulator might be used for a preliminary assessment to ensure that an individual is ready to make the best use of an opportunity for assessment on real equipment. As an alternative, a preliminary assessment on a simulator might be augmented later by a more limited assessment on real equipment. Additionally, simulators provide an opportunity for assessment under relatively infrequent, high-risk conditions. Given these alternatives, a decision would have to be made as to whether the greater effectiveness of assessment on a more costly simulator or on real equipment justifies the increased cost. We have proposed a systematic method for evaluating simulators and must leave it to others in the maritime industry to design a broader program of performance-based assessments that benefits from the capabilities of simulators. A parallel effort to ours, to systemically identify the features needed by engine room simulators to support mariner assessment, reached a similar conclusion that simulators need to be incorporated into a broader program of mariner assessment (Stutman, 1999).

## References

1. International Maritime Organization (1998), "Guidance on Shipboard Assessments of Proficiency," *MSC/Circ. 853*. London, Author.
2. International Maritime Organization (1979), "Performance Standards for Automatic Radar Plotting Aids (ARPA)," *Resolution No. A.422XI*. London, Author.
3. International Maritime Organizations (1971), "Performance Standards for Navigational Radar Equipment," *Resolution No. A.222VII*. London, Author.

4. International Maritime Organization (1995), "Seafarer's Training, Certification, and Watchkeeping (STCW) Code." London, Author.
5. McCallum, M.C., A.M. Forsythe, A. Barnes, M.W. Smith, J. Macaulay, G.R. Sandberg, J. Murphy, and P. Jackson (1999), "A Method for Developing Mariner Assessments," (Report UDI Number R&DC 202-99). U.S. Coast Guard Research & Development Center. Groton, CT. In revision.
6. McCallum, M.C., A.M. Forsythe, J.M. Nunnenkamp, G.R. Sandberg, and M.W. Smith (1999), "Developing Performance-Based Assessments of Mariner Proficiency," (Report UDI Number R&DC 303-99). U.S. Coast Guard Research and Development Center. Groton CT. In revision.
7. Raby, M., A.M. Forsythe, M.C. McCallum, and M.W. Smith (1999), "Evaluating Simulators for Assessments of Mariner Proficiency," (Report UDI Number R&DC 150-99). U.S. Coast Guard Research and Development Center. Groton, CT. In revision.
8. Stutman, P. (1999), "Comparing the Assessment of Marine Engineers Using Engine Room Simulators Versus Shipboard Practical Demonstrations," *Proceedings of ICERS/4: The Fourth International Conference on Engine Room Simulators*.